

DOCUMENT RESUME

ED 235 828

IR 050 486

AUTHOR Matsumura, Tamiko
TITLE Science Information System in Japan. NIER Occasional Paper 02/83.
INSTITUTION National Inst. for Educational Research, Tokyo (Japan).
PUB DATE Mar 83
NOTE 21p.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Cataloging; Check Lists; *Databases; Design Requirements; Foreign Countries; *Information Centers; *Information Networks; Information Retrieval; *Information Services; Interlibrary Loans; *National Programs; Online Systems; Program Design; Union Catalogs

IDENTIFIERS Information Transfer; *Japan; Resource Sharing

ABSTRACT

This paper describes the development of a proposed Japanese Science Information System (SIS), a nationwide network of research and academic libraries, large-scale computer centers, national research institutes, and other organizations, to be formed for the purpose of sharing information and resources in the natural sciences, technology, the humanities, and the social sciences. It is noted that SIS will include sharing of primary documents, secondary information retrieval, provision of library holdings information, machine-readable library cataloging, and production of specialized databases. The paper is divided into nine sections covering: (1) background information; (2) SIS basic concepts; (3) SIS functions; (4) the structure of SIS, which is to include a Science Information Center (SIC) with planning, coordinating, service, research and development, and educational/training functions; (5) proposed SIC activities, service provision methods, and users, with a chart outlining an SIC development schedule; (6) the SIC database system, with a list of general, information retrieval, cataloging, and other system requirements; (7) the performance design of the SIC database system, with an outline of hardware, capacity, and access requirements in three stages of SIC development; (8) the SIS network system, which will include information retrieval, library cataloging and holdings information, and document delivery networks; and (9) the significance of SIS in information transfer in Japan. (ESR)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED235828

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ↓ This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

SCIENCE INFORMATION SYSTEM IN JAPAN

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Toshio Kanaya

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Section for Educational Co-operation in Asia
NATIONAL INSTITUTE FOR EDUCATIONAL RESEARCH OF JAPAN
5-22 Shimomeguro 6-chome, Meguro-ku, Tokyo.
March 1983

IR050486

N O T E

This paper was prepared by Professor Tamiko Matsumura at the University of Library and Information Science for dissemination under NIER's Information Services Programme among Unesco Member States in Asia and the Pacific.

I. BACKGROUND

In October 1973, the Science Council, an advisory body to the Minister of Education, Science and Culture, published a report entitled The Immediate Basic Policies for Science Development. The Report described the state-of-the-art of science in Japan, identified the problems and made recommendations for future policies with regard to the promotion of science. Although the Report was primarily concerned with science, it made specific reference to science information, indicating a need for a systematic approach to the improvement of information transfer in the academic community, and set out the basic policies necessary to achieve this goal.

In view of the recommendations in this Report, the Ministry of Education, Science and Culture through its grants-in-aid scheme supported a number of research projects to develop expertise with information and library systems, to name only a few, information handling in specific subject fields, machine-readable cataloguing and the production of a serials union list data base.

In April 1978, the Science Council appointed a task force group, the Subcommittee on Science Information, within the Committee on Science Information and Materials, to devise an optimum scheme for improving the transfer of information to meet the needs of science research. In November 1978, the Minister of Education, Science and Culture gave an official assignment to the Science Council to study what form a future information system should take in Japan. An intensive study was carried on by the Subcommittee for twelve months and in June 1979 the Science Council published an interim report on a new science information system with a set of recommendations in order to ask for opinions and comments from various sectors of society. After giving due consideration to the opinions and comments presented, the Science Council submitted its final report to the Minister in January 1980.

The Report, entitled A New Plan for a Science Information System, described the basic concept of the proposed Science Information System, identified the major functions, activities and structure of the system, and made a set of recommendations including the need for the establishment of a Science Information Centre which was to play a key role in this new system. After its publication the Report received many positive responses from the academic community asking for its early implementation.

In response to such expressions of support, the Ministry of Education, Science and Culture set up in May 1980 a committee to investigate the organization, activities and facilities, including a computer system, of the Science Information Centre (SIC). The study was carried out by two subcommittees: one on the technical

aspects of a computer system of the Centre, and the other on the information and library network. During the year 1981 further studies were continued on the basic requirements of the data base system of the Centre and the system development started later in the same year.

The present paper describes the basic concept, activities, structure and network of the proposed Science Information System and the data base system of the Science Information Centre, which will play a key role in the System, based on the studies undertaken prior to March 1982.

II. BASIC CONCEPTS OF THE SCIENCE INFORMATION SYSTEM

The Science Information System (SIS) is based on the concept of sharing resources, human as well as material, in all fields of knowledge in order to support research activities in the academic community of the country. It covers, therefore, the humanities and social sciences as well as the natural sciences and technology.

It is an inter-university information and library network whereby all existing library and information systems and services together with computer centres in universities and research facilities are linked together with proper coordination and co-operation into a dynamic nation-wide network.

It is a comprehensive system covering secondary information retrieval, the provision of primary documents, information on library holdings and machine-readable library cataloguing.

III. FUNCTIONS OF THE SIS

The Science Information System covers a wide range of activities, which may conveniently be divided into three major areas: first, the provision of primary information, secondly, information retrieval services, and thirdly, the production of data bases.

1. Provision of Primary Information

It was recommended in the 1980 Report that the existing primary information resources be strengthened rather than establishing a new central resource to provide a nation-wide service. The university library collections together constitute a major documentary resource in the country. However they have not been sufficiently exploited, and also need to be strengthened and improved with a view to ensuring a balanced coverage of subjects, languages, and other aspects.

Within this frameworks, a resource centre library scheme has been initiated, whereby university libraries with strong collections in certain subjects, are designated as centre libraries, which collect primary literature published elsewhere in the respective subjects, and provide an interlibrary loan service throughout the country. Technical reports, proceedings and other similar types of materials should be handled by research institutions in specialized fields which have more ready access to these materials. Historical documents, government publications are rather different from those acquired from ordinary circulation channels, therefore they should be handled separately.

Equally as important as the strengthening of resources is the improvement in the dissemination of information. In this respect all libraries and organizations in the SIS are expected to make a positive effort to make their collections readily available for general use and to promote inter-institutional co-operation in the supply of documents.

2. Provision of Secondary Information Retrieval System

Several information retrieval systems have been developed in universities mainly as research experiments providing certain services to academic users. However these services are rather an extension of research activity and are limited in the scope of the service, the kind of data base offered and the system capability. Experience with these services, however, contributed to making researchers aware of the need for secondary information for their research.

The secondary information retrieval service of the SIS covers all subject fields with major data bases available elsewhere which are necessary for research, and provides nation-wide retrieval service through the network. The computer network which already exists among large-scale computer centres will be expanded to form a national network for a data base service. In this system the general computing power and the hardware/software facilities can also be shared throughout the country.

3. Data Base Production

The kind of data base which will be produced in the framework of the SIS will contain rather specialized and highly advanced information produced as the result of research. Such data bases are not likely to lend themselves to commercial interests and their production will probably take place where research is conducted or where close contact with researchers can be obtained, for instance, in a national research institution or within the framework of a learned society. Potential areas for data base production may include fields where Japan has a significant accomplishment in research but where the results have not been properly disseminated

outside the country and fields where Japan can contribute to the international co-operation in the field of data base production and information transfer.

In certain fields data base production is totally dependent on the voluntary efforts of individual researchers, and some measures should be taken to expand and strengthen the current grants-in-aid scheme to support existing activity and to promote future data base production.

Due consideration should be given to the quality control of input data and also to the evaluation of numerical and other data.

IV. STRUCTURE OF THE SIS

The SIS is a network based on the concept of sharing both human and material resources in the academic community. The constituent members of this network include libraries in universities and research institutions, large-scale computer centres, national research institutes and other organizations.

(1) Large-scale computer centres, national research institutions, and research institutes in universities will have, in addition to their conventional activities, the following functions in the SIS:

- 1) To serve as the nodes in the science information network,
- 2) To be responsible for the production of data bases, or to provide facilities for the production particularly of graphic or numerical information, and to maintain such data bases for search services, and
- 3) To co-operate with the central unit of the SIS in research and development.

(2) Functions shared by other computer centres in universities and research facilities include:

- 1) To provide facilities for the development of data bases indigenous to the country, and
- 2) To co-operate with the central unit of the SIS in research and development.

(3) University libraries will share important responsibilities in the SIS as the front-end and back-end of the system and should assume the following functions:

- 1) To serve as a window or terminal for information retrieval services,

- 2) To provide primary information from their own collections, and to serve as a liaison with resource centre libraries and other libraries in an interlibrary loan service, and
- 3) To contribute to building a national data bank.

In addition to those functions assumed by constituent organizations, certain mechanisms will be required in order to ensure that the varied components in the network function effectively in coordination and co-operation. For this purpose it was recommended that a centre for the SIS should be established which would be provided with the following functions.

(4) The functions of the Science Information Centre (hereafter abbreviated as SIC) may be categorized into the following areas:

1) Planning and coordination function

- a. To plan the appropriate distribution of data bases within the SIS for effective service.
- b. To work out feasible future plans and to monitor their implementation (for example: review current activities and plan future programmes in information retrieval, data base production, authorization of research files, and education and training).
- c. To promote the standardization of information processing and handling with a view to ensuring integrity and conformity in the network.
- d. To coordinate with external organizations in and outside the country.

2) Service function

- a. To provide a secondary information retrieval service with data bases which it is considered can be more effectively processed and maintained centrally.
- b. To support library cataloguing and provide information on holdings of primary documents available in the country.
- c. To provide a clearing service concerning data bases distributed throughout the network, research activities taking place in the system and other matters related to the SIS.

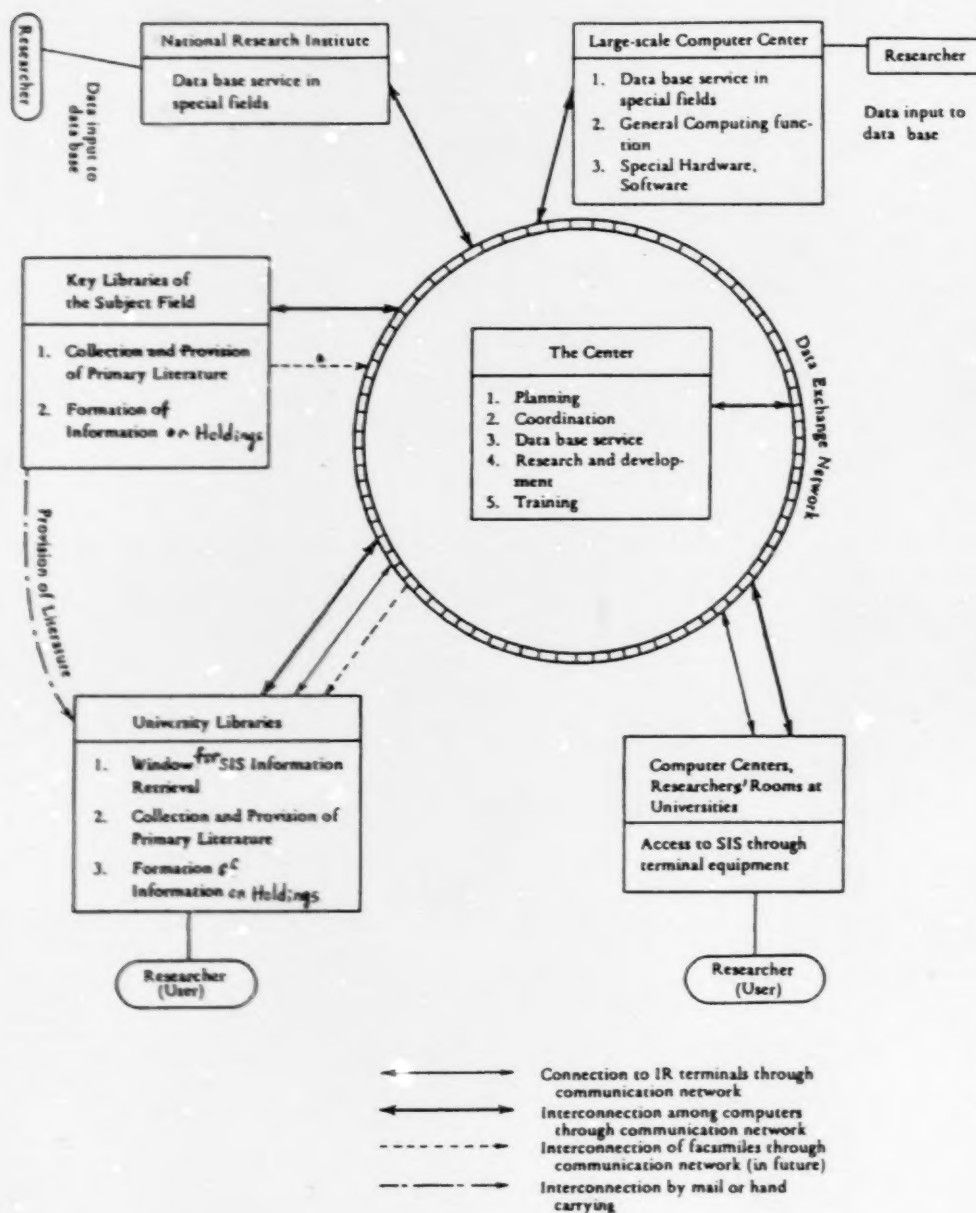
3) Research and development function

To conduct research and development related to the operation of the SIS.

4) Education and training function

To provide education and training facilities to both users and staff of the SIS.

The following chart shows the conceptual structure of the SIS with functions shared by the constituent organizations.



IV. ACTIVITIES OF THE SCIENCE INFORMATION CENTRE (SIC)

The Science Information Centre is to play a key role in the information and resource-sharing network system. The proposed activities of the SIC are summarized in the following table:

Activities	Contents of activities	Provision method	Users
Machine-readable library cataloguing	1) Provision of MARC-stored bibliographic data on books and journals in the humanities, social sciences and natural sciences. 2) Production of individual library catalogue.	On-line interactive mode through a new data exchange network.	National, public and private university libraries.
Location information data base generation and retrieval service.	1) Generation of location information data bases. 2) Location information retrieval service.	On-line interactive mode through a new data exchange network.	National and private university libraries, other libraries.
Provision of primary information	Literature in the humanities, social sciences and natural sciences.	Provision of photocopies or loans through university libraries.	1) Researchers in national, public, and private universities. 2) Users in the private sector.
Generation and provision of domestic data bases.	1) Production of science information data bases in fields where no data bases are available and their provision to domestic and overseas researchers. Input of information evaluated by researchers in a format and language appropriate for international dissemination. 2) Co-operative input into international data bases.	1) On-line interactive mode through a new data exchange network. 2) Provision of data bases.	1) Researchers in national, public, and private universities. 2) Domestic and overseas data base service organizations.

(To be continued)

Activities	Contents of activities	Provision method	Users
Research and development.	R&D necessary for the operations of the SIC and the SIS.	Reports at meetings or in publications.	Researchers in universities.
Provision of secondary data base information.	1) Provision of needed information from data bases in the humanities, social sciences and natural sciences. 2) Provision of numerical or graphic information in particular fields (such as geomagnetism, etc.) from data bases.	On-line interactive mode through a new data exchange network.	Researchers in national, public, and private universities.

The development schedule for SIC activities is proposed as follows:

Year	Secondary information retrieval service.	Cataloguing and location information.
-3	Start designing data base management system. Start designing network management.	Start system analysis for machine-readable cataloguing.
-2	Design user registration system. Completion of data base management system, network control.	Design cataloguing system. Design housekeeping language.
-1	Start loading data bases onto the computer. Start trial provision of secondary information retrieval service. Start trial of union catalogue information service	Start trial of machine-readable cataloguing and location-information in respect of specified libraries. Start MARC retrieval service.
0	Start full-scale data base service.	Trial production of union catalogue.
+1		Start shared-cataloguing service.

V. DATA BASE SYSTEM OF THE SIC

1. Common Requirements

The Centre's data base system consists of two parts, each with its own functional characteristics, but the basic requirements of the data base system are common to both the retrieval subsystem and the catalogue subsystem. The main features of the data base system include:

(1) Intersectional use of data bases

This is a function whereby more than one data base may be linked together in the course of a search. There are a number of conceivable means of achieving this function depending on the type of data base to be searched and the mode of searching. The functions which should be operational in the first phase of development include:

- 1) A search question to one data base and its result to be available for use when another data base is searched.
- 2) Ability to search more than one data base at the same time.
- 3) Functional linking together of the basic data bases designated by the SIS so that they look to users like a single data base.

In addition, the function whereby users can link a number of data bases to form a single logical data base is expected to be operational at a later stage of development.

(2) Other requirements

- 1) Data base management function, for example, flexibility to respond to the growth and change of individual data bases and to maintain independence and the safety of data.
- 2) Capability to process Japanese, Roman, Cyrillic and other characters, and a user interface with which various data bases may be searched by means of a single retrieval language.
- 3) Function of supporting users by providing instruction for use and processing claims on-line.
- 4) Function of supporting terminals for graphic information as well as dedicated terminals for the catalogue system.

2. Retrieval System Requirements

In addition to the common requirements, the retrieval system should be able to meet the following specific requirements:

(1) A search output should be able to be appropriately edited and filed for storage and possibly input later again as a search question.

(2) A plural form to be identifiable with its singular form.

(3) A single data base should be able to be divided into a number of sub-data bases according to the conditions set forth by a user and able to be further linked for a collective search.

(4) A system message should be able to be output in both English and Japanese but a user may choose either of the two.

(5) The types of terminals able to be used in the retrieval system to include:

- 1) Input/output of Chinese characters, Kana alphabets, Roman alphabet and numerals (Japanese language terminal).
- 2) Output as in 1) and input of Kana alphabets, Roman alphabet and numerals.
- 3) Input/output of Kana alphabets, Roman alphabet and numerals.
- 4) Input/output of only Roman alphabet and numerals.

Regarding the graphic function and the page mode editing function, connection to the system may be restricted to the terminals which meet certain standards.

3. Catalogue System Requirements

- (1) Support of library automation

In order to provide proper support for the automation of library activities, there is a need for a hierarchical network in which the SIC, the central libraries of several geographical regions (hereafter called 'regional centres') and individual libraries will be functionally linked for co-operation and coordination. Some features of such a network include:

- 1) Cataloguing of books and journals will be done on-line with dedicated terminals in individual libraries.
- 2) The catalogue data input from terminals to the Centre will

be made available to all libraries in the network through the retrieval system.

- 3) The Centre will process interlibrary loan transactions.
- 4) The Centre will provide bibliographic information from its data bases to support automation of other library activities. The question of whether individual libraries in the network should be directly linked to the Centre or linked through a regional centre library should be left with individual libraries depending on their approach to library automation, geographical environment associated with the telecommunication network and various other factors.

(2) General features of the catalogue system

The catalogue system will have two subsystems: a book cataloguing system and journal cataloguing system.

1) Book cataloguing system

Authority files are necessary in order to control the quality of bibliographic data input from terminals in libraries. In the initial phase of operation, authority files available elsewhere such as the Name Authority File of the US Library of Congress will be used, and a further study will be made on the production of locally based authority records generated in the network. In order to fill the time-lag between the publication of a book and its catalogue record being made available in MARC, a catalogue record pending system will be provided whereby only the code of a book will be input at the time of cataloguing to be replaced later by full bibliographic data. The output of catalogue cards and book labels will be done basically at terminals in individual libraries. However, the Centre may provide an output service for a book-form and other forms of catalogue and for those containing characters such as non-standard Chinese characters which are not processed by ordinary terminals in libraries. As regards the future form of a library catalogue, further studies will be needed.

2) Journal cataloguing system

The Centre also provides LC-MARC (Series) and other authority files for quality control of the system. The input of bibliographic data will be done on-line from the terminals in libraries. Before the system is made operational, the Centre will have built up a data base based on the machine

readable Union List of Scientific Periodicals, which contains bibliographic data of scientific periodicals in all subject fields together with the records of holdings of all university libraries in the country. The input of data concerned with daily acquisition in individual libraries will be done in remote batch processing.

3) Interlibrary loan system

The interlibrary loan transactions which are currently performed by ordinary mail will be speeded up by the Centre's electronic mail service. Data communication will be conducted in a prescribed telegraphic message, and the charge accounting and various statistics will also be processed by the same system.

4) Technical problems in the catalogue system

Since the cataloguing system is basically related to the retrieval system, it would seem logical to have a common data base management system for both systems. The major problems inherent in the cataloguing system lie in the treatment of uncoded Chinese characters (characters other than those coded by JIS and other standards) and those characters which have variant written forms. Although the Centre system should eventually manage to process these characters, further studies will be required on whether it will be feasible for the system to handle all potentially uncoded characters and possible variant forms, and how the work loads will be divided between the Centre and the terminals. These will influence the specifications of dedicated terminals, although it has proved to be evident at this stage that terminals for cataloguing should be provided with the American Library Association character set and 8,000 to 10,000 Chinese characters (JIS Nos. 1 and 2 standards and their extension) in addition to the Roman alphabet, Kana alphabets and numerals as minimum requirements.

4. Data Base Service Management

The management functions associated with the Centre's services may roughly be classified as follows:

- (1) Updating and storage of secondary information data bases.
- (2) Centre system maintenance.

(3) Management of data base use (primarily control of users and charges).

- (4) Provision of guidance, education and training.
- (5) Other management activities.

The updating of secondary information data bases is carried out on a daily basis. The potential users of the Centre system may well exceed 15,000 (excluding libraries and other organizational users), which is about 10 percent of individuals legitimately able to use the services. In view of the enormous amount of work involved in these management activities, an optimum system will be required for the efficient provision of services and the effective operation of the system. Due consideration will also have to be given to precautions against breakdowns and disasters.

5. Network

As regards the networking between the Centre and the terminals, it is conceivable that either the N-1 protocol with which the large-scale computer centres in universities are currently linked (hereafter called "N-1") or a general purpose protocol will be used. However a general-purpose protocol is hardly suitable for the Centre system whose functions surpass ordinary data processing (for example, page mode editing and the processing of uncoded characters) and which requires local protocol processing. In order also that the various terminals which are currently linked to these large-scale computer centres should be connected with the Centre without radical change, the N-1-based protocol seems to be optimum for the system. However, certain additional functions (for instance, file transmitting function) will be required to meet the needs of the Centre system.

6. Other Requirements

There is a need to support research and development for the improvement of the Centre system and services. To give one example, there is a need for the availability of a system dedicated to research and development with the same operating characteristics as the Centre's host computer and which can be used for testing software, newly introduced hardware and data bases and new communication technologies.

VI. PERFORMANCE DESIGN OF THE SIC DATA BASE SYSTEM

A performance analysis of the data base system was conducted on the basis of the volume of information dealt with at the Centre, the number of terminals connected to the Centre and simultaneously in operation, and other predictable conditions.

1. Premises

The foreseeable conditions of performance are as follows:

(1) Type and volume of information processed at the Centre

Type of information	Number of records	Volume
Secondary information	52M	208GB
Catalogue information		
Bibliographic information	10.2	21
Location information	90	9
Name authority information	2	0.4
Dictionary information	1.2	0.5
(conversion of Kana alphabets to Chinese characters, etc.)		
		(Total 239)

(2) Number of terminals in simultaneous operation

Type of terminal	Number of units
Terminals for retrieval	100 - 300
Terminals for cataloguing	350 - 1,000

(3) Volume of communication (in the initial phase of development)

Type of service	Average number (times/a day)	Volume of communication (packets/a time)	Volume of communication (M packets/a day)
Information retrieval (secondary and catalogue information)	1,000	900	0.9
Input of catalogue information			
Monographs	28,000	50.8	1.4
Journals	3,800	100	0.38

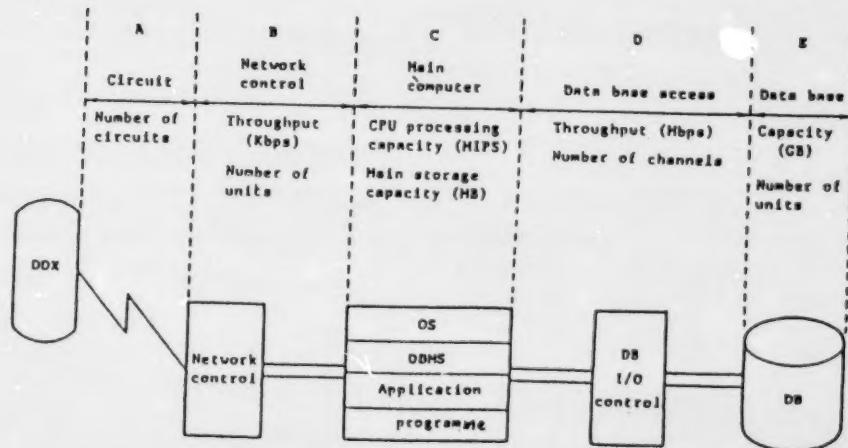
(4) Type of processing and response time

- a. Retrieval system will be operated in TSS mode and the catalogue system in real-time processing.
- b. The target value of response time of the catalogue system will be three seconds, of which one second will be used for the network processing.

2. Findings of Performance Analysis

(1) Components of the system

Based on the foregoing premises, the SIC system will be as follows:



(2) The findings of the performance analysis are as follows:

Phase	Number of simultaneous access terminals		A Circuit	B Network control		C Main computer	
	Catalogue	Retrieval	Number of circuits	Number of units	Through-put (Kbps)	CPU processing capacity (MIPS)	Main storage capacity (MB)
1	350	100	6	3	288	32	67
2	700	200	7	7	336	60	118
3	1,000	300	10	10	480	87	165
Condi- tions			Data trans- fer rate 48 Mbps	FEP processing capacity is hypothesized at 2 MIPS.		*On-line MIPS *CPU load rate 70%	

Phase	D Data base access				E Data base			
	Disk		Drum		Disk		Drum	
	Number of channels	Through-put (Mbps)	Number of channels	Through-put (Mbps)	Number of units (spindles)	Capacity (GB)	Number of units	Capacity (GB)
1	3	9	4	6	413	248	15	0.23
2	5	15	7	11	429	258	19	0.29
3	7	21	10	15	445	267	23	0.36
Condi- tions	Data transfer rate 3 Mbps		Data transfer rate 1.5 Mbps		Capacity 600 MB/spindle		Capacity 15 MB/drum	

- Notes: 1) Phases 1-3 represent the stage of development.
2) Figures for the system management are not included.
3) In the first phase, one network control unit processes two circuits.

VII. NETWORK SYSTEM OF THE SIS

The network of the SIS will be divided functionally into two parts: one is the information retrieval network for data bases (bibliographic, numerical, graphic and others), and the other is the network for machine-readable cataloguing and information on holdings.

1. Information Retrieval Network

The organizations which will take part in the information retrieval services of the SIS include large-scale computer centres, research institutions provided with large computer facilities as well as the SIC itself which will play a key role in this network. Relatively simple terminals will be installed in libraries and offices of individual researchers for direct access to data bases.

2. Library Cataloguing and Holdings Information Network

The libraries in universities and research institutions will play a major role in this network. MARC data bases and other bibliographic sources as well as library holdings data bases will be located at the Centre. It is anticipated that the structure of the network will be hierarchical: terminals in individual libraries will be connected to a regional center library serving as the processing centre for specific regions, which will be linked with the Science Information Centre. Considering the rapid progress in information technology, however, the optimum network system will be left for continued study.

3. Document Delivery Network

The location information on documents will be acquired through the information retrieval network from the data bases of the SIC and an interlibrary loan request will be sent automatically to the library which holds the requested item. Regarding the mailing of materials a study is being undertaken on the feasibility of electronic transmission through this network.

VIII. SIGNIFICANCE OF THE SIS IN INFORMATION TRANSFER IN JAPAN

Science research which takes place in university and research institutions encompasses all fields of knowledge and constitutes the major part of the basic research activity in the country. Those who are engaged in research are users of information and at the same time the generators of information. The SIS is characterized by its close relations with the academic community where information is generated and utilized. It is also characteristic that the system is a

comprehensive information- and resource-sharing network. Although the SIS was initially conceived and planned in response to the needs of the academic community, it intends to extend services to other sectors of society through coordination and co-operation with other information systems and services. It is expected that the SIS will contribute to the development of an optimum system for information transfer in Japan, thus making a contribution also to the world-wide transfer of science information.